DOCUMENT RESUME

ED 218 098 SE 037 898

AUTHOR Bedian, Vahe
TITLE Evaluation of Diagnostic Tests and Decision Analysis. Applications of Probability and Statistics to Medicine. Modules and Monographs in Undergraduate Mathematics and Its Applications. UMAP Module 377.
INSTITUTION Education Development Center, Inc., Newton, Mass.
SPONS AGENCY National Science Foundation, Washington, D.C.
PUB DATE 79
GRANT SED-76-19615-A02
NOTE 47p.

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Answer Keys; *College Mathematics; Higher Education; Instructional Materials; *Learning Modules; *Mathematical Applications; *Medical Education; Medicine; *Probability; Problem Solving; *Statistics; Supplementary Reading Materials

ABSTRACT

The material looks at: 1) Choosing a Clinical Test; 2) Evaluation of Diagnostic Tests; 3) Multi-Disease and Multi-Test Analysis; and 4) Medical Decision Analysis. It is felt users will be able to: 1) calculate the predictive value of a positive or negative test result in model clinical situations; 2) estimate the sensitivity and specificity required from a test for a desired certainty of diagnosis; 3) analyze alternative modes of using multiple tests and choose the mode that optimizes the information gained about a single disease or group of diseases; 4) describe a clinical situation in terms of decision points and probabilistic events connected in a decision tree; and 5) evaluate overall strategies of diagnosis and treatment in terms of probabilities, costs, and benefits of outcomes. Exercises and a sample exam are included, with answers to both provided. (MP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED218098

UMAP

MODULES AND MONOGRAPHS IN UNDERGRADUATE MATHEMATICS AND ITS APPLICATIONS

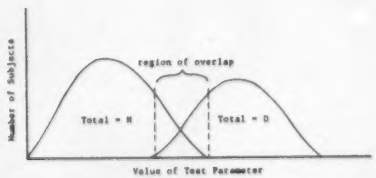
Birkhäuser Boston Inc.
380 Green Street
Cambridge, MA 02139

MODULE

377

Evaluation of Diagnostic Tests and Decision Analysis

Vahe Bedian



Applications of Probability and Statistics to Medicine

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

National Science Foundation

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EVALUATION OF DIAGNOSTIC TESTS AND DECISION ANALYSIS

by

Vahe Bedian
Department of Biophysical Sciences
State University of New York at Buffalo
Buffalo, NY 14214

TABLE OF CONTENTS

1. CHOOSING A CLINICAL TEST 18

2. EVALUATION OF DIAGNOSTIC TESTS 19

3. MULTI-DISEASE AND MULTI-TEST ANALYSIS 28

4. MEDICAL DECISION ANALYSIS 38

5. EXERCISES 45

6. SAMPLE EXAM 48

7. ANSWERS TO EXERCISES 50

8. ANSWERS TO SAMPLE EXAM 56

Intermodular Description Sheet: UMAP Unit 377

Title: EVALUATION OF DIAGNOSTIC TESTS AND DECISION ANALYSIS

Author: Vahe Bedian
Department of Biophysical Sciences
State University of New York at Buffalo
Buffalo, NY 14214

Review Stage/Date: III 7/27/79

Classification: APPL PROB & STAT/MED

Prerequisite Skills:
Basic concepts of probability theory: probability, conditional probability, statistical independence and multiplication rule, mutually exclusive states, and expectation values.

Output Skills:
Students using this module will be able to:
1. Calculate the predictive value of a positive or negative test result in model clinical situations.
2. Estimate the sensitivity and specificity required from a test for a desired certainty of diagnosis.
3. Analyze alternative modes of using multiple tests and choose the mode that optimizes the information gained about a single disease or group of diseases.
4. Describe a clinical situation in terms of decision points and probabilistic events connected in a decision tree.
5. Evaluate overall strategies of diagnosis and treatment in terms of probabilities, costs, and benefits of outcomes.

The Project would like to thank Jay E. Anderson of Idaho State University, Earl Faulkner of Brigham Young University and member of the UMAP Statistics Panel, and members of the UMAP Biomedical Sciences Panel for their reviews, and all others who assisted in the production of this unit.

This material was prepared with the partial support of National Science Foundation Grant No. SED76-19615 A02. Recommendations expressed are those of the author and do not necessarily reflect the views of the NSF or the copyright holder.

© 1979 EDC/Project UMAP
All rights reserved.

1. CHOOSING A CLINICAL TEST

1.1 What Makes a Good Diagnostic Test?

An ideal clinical test would tell the physician with absolute certainty whether or not the patient has a suspected disease without any risk or cost to the patient, and would provide prognostic information in case of an abnormal finding. However, most tests expose the patient to some risk, as well as to financial and psychological stress, which, when combined, constitute the overall cost of the test. Diagnostic tests also yield results with a limited level of certainty, and a positive result may tell little about subsequent development of the condition. Two questions, therefore, arise in considering any diagnostic procedure: what is the level of certainty associated with a positive or a negative finding? What is the cost of this procedure and how does it compare with the expected benefits (i.e., lower mortality or morbidity through earlier or better treatment) to the patient? Approaches for evaluating diagnostic procedures will be developed in Sections 1 through 3. Whether a procedure should be adopted, and the cost/benefit considerations involved, will be discussed in the context of decision theory in Section 4.

Here are some important considerations in choosing a clinical test:

Relevance: does the test measure a parameter that is clearly related to the disease under consideration?
Significance: will the test result change the strategy of the physician or eliminate some alternatives from consideration?
Usefulness: does the test have the potential to increase the level of certainty with which a diagnosis has already been made?

The first two conditions are prerequisites for any diagnostic procedure. While many clinical tests may satisfy these two conditions, the physician has to choose the test with the highest usefulness and the least cost and risk to the patient.

The usefulness of a diagnostic test depends on how much more certain we can be about the presence (or absence) of a disease given that the test result is positive (or negative), compared to the level of certainty with which the diagnosis was made prior to the test. The higher this increment in the level of certainty, the more useful the test is.

1.2 When is a Probabilistic Approach Necessary?

When the diagnostic procedure under consideration is obviously safe and cheap, and the expected results are obviously useful, the physician may adopt the procedure without a quantitative analysis. An example of such a situation may be: taking the blood pressure and heart rate of a patient complaining of chest pain. The physician's initial suspicion of a heart disorder may have been only 5-percent certain, and a positive finding (abnormal heart rate and/or blood pressure) would definitely raise this level of certainty and lead to further exploration.

If the diagnostic procedure involves substantial risk or high cost to the patient, or when it is not obvious that results will be useful, an explicit probabilistic analysis of the procedure is called for. An example of this situation may be: consideration of a needle biopsy of the liver when previous tests indicate the presence of a malignant tumor with 70-percent certainty. In this situation, it is imperative to ascertain that the biopsy would be useful; i.e., a positive result would indicate the presence of a malignant tumor with more than 70-percent certainty. Even if this condition is satisfied, the risk and cost of the procedure would have to be compared with the potential benefits of the additional information that can be obtained from the test. The test may be indicated when the previous information ascertained the presence of a tumor with 70-percent certainty but may be unjustified when the a priori certainty was much lower and other less dangerous preliminary tests were available or when the diagnosis has already been established with certainty and the biopsy cannot add significantly to the information already available.

Sections 2 and 3 will be concerned with estimating the level of certainty (predictive value) of a positive or negative finding in a clinical test, given the level of certainty of initial diagnosis (prevalence) and statistical information about the effectiveness of the test as a diagnostic tool.

2. EVALUATION OF DIAGNOSTIC TESTS

2.1 Sources of Variation

A population is defined by identifying the characteristics common to all its members; e.g., all males between the ages of 20 and 65 that live in the United States and smoke more than 10 cigarettes per day. A sample of a population is a representative group of its members. The sample is studied with the hope of gaining insight into characteristics of the whole population.

Measurement of a particular anatomical, physiological, or biochemical parameter in a sample of a population always shows a distribution of values. Two sources of variation contribute to the diversity of values obtained: biological and methodological. Many factors contribute to biological variation, including age, sex, race, nutrition, diurnal changes, general state of health, in addition to the presence of a particular disease. Methodological errors can affect both the precision and accuracy of results. Precision refers to repeatability of a measurement, while accuracy refers to how close the obtained value is to the true value. Statistical or random instrumental errors affect precision; e.g., a spectrophotometer used for repeated measurements of the concentration of a metabolite in the same sample may yield a distribution of values around a mean. Systematic errors affect the accuracy of the result, and may be inherent to the approach or introduced by inadequate procedures. For instance, a clinical laboratory may report consistently low values for a blood constituent because samples are not refrigerated adequately.

All these factors contribute to the difficulty of using tests for diagnostic purposes. In spite of these limitations, we can make inferences about the significance of a test result, given statistics about the test and the disease in question. This is an attempt to go a step beyond the concept of "normal range," which has several different statistical interpretations and tells little about certainty of diagnosis when a particular test result is obtained.

2.2 Statistics of A Clinical Test

When a new clinical test is developed, it is administered to a population comprised of subjects that are known to be free of the disease tested for (H), and subjects that are known to have the disease (D). In this way, two distribution curves for the value of the measured parameter are obtained. For most clinical tests there is a region of overlap between these two curves where it is impossible to distinguish between results from healthy and diseased subjects (Figure 1).

A test result is considered positive or negative depending on its position with respect to an arbitrary demarcation line or referent value. The choice of a demarcation line defines four populations related to four classes of test results (Figure 2):

True positives (TP): diseased subjects with a positive result.

False negatives (FN): diseased subjects with a negative result.
True negatives (TN): healthy subjects with a negative result.
False positives (FP): healthy subjects with a positive result.

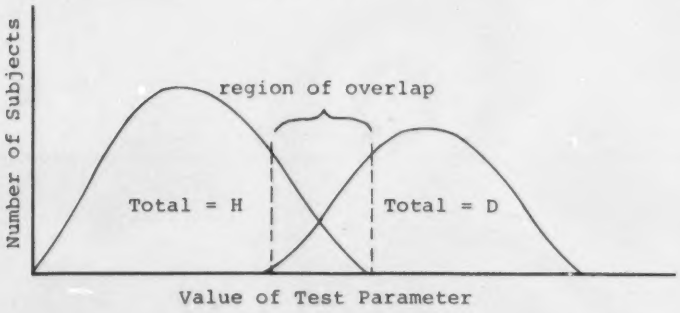


Figure 1. Distribution curves of test parameter for healthy and diseased populations. (Both axes are in arbitrary units.)

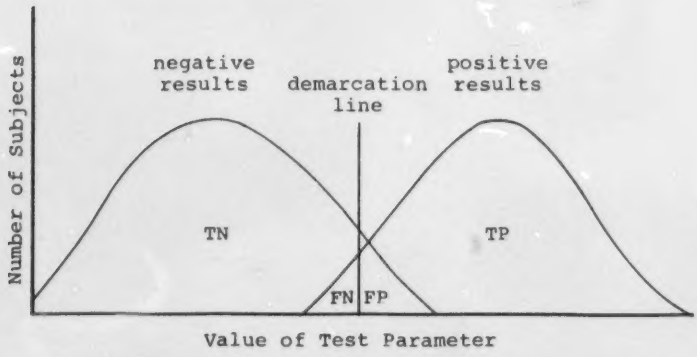


Figure 2. Four populations defined by a demarcation line and two distribution curves.

False negatives and false positives are also referred to in the literature as type I and type II, α and β, or omission and commission errors, respectively.

The total number of subjects in each of the two populations is given by:

(1) $D = TP + FN.$

(2) $H = TN + FP.$

The application of any clinical test implies the choice of a demarcation line resulting in the four populations shown in Figure 2. A diagnostic test is then characterized by two quantities:

Sensitivity (s): fraction of diseased population with a positive result.

Specificity (c): fraction of healthy population with a negative result.

In terms of TP, FN, TN, FP:

(3) $$s = TP/(TP + FN) = TP/D.$$

(4) $$c = TN/(TN + FP) = TN/H.$$

A good clinical test has a small region of overlap between the two distribution curves, resulting in low numbers of false results and high values of sensitivity and specificity. The demarcation line can be moved to give high sensitivity, but this will result in lower specificity. Conversely, high specificity can be obtained at the price of lower sensitivity.

Along with sensitivity and specificity, another statistical parameter is frequently quoted in descriptions of clinical tests. This is the efficiency (e) of the test defined as:

$$\text{Efficiency} = \frac{\text{Total number of correct results}}{\text{Total number of results}}$$

(5) $$e = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{H + D}.$$

While the sensitivity (s) and specificity (c) are both essential for estimating the four populations TP, TN, FP, and FN, the efficiency (e), which pools them all together, is generally less useful for evaluation of clinical tests, since it will not distinguish between a test of high sensitivity and low specificity, and one of high specificity and low sensitivity.

2.3 Populations and Probabilities: Two Parallel Interpretations

The discussion above has been in terms of numbers of subjects in certain populations and subpopulations; it, therefore, assumes a statistically large number of subjects in these populations. This is the situation we encounter in epidemiological studies. The same criteria apply, however, when the test is used for diagnosing an individual

patient. By dividing the subpopulations TP, FN, FP, TN by the total population (Tot)

(6) $$Tot = D + H,$$

we obtain four fractional quantities that can be interpreted as the probability of obtaining a particular type of result for a given individual. Thus, TN/Tot would be the probability of obtaining a true negative result.

The ratio D/Tot deserves special attention, since it is defined to be the prevalence (p) of a disease:

(7) $$p = \frac{D}{Tot} = \frac{D}{D + H}.$$

The epidemiological interpretation of prevalence is "the fraction of the total population that has the disease at a given time." Thus, it is an important statistical measure of the occurrence of the disease in a given population.

In the case of an individual patient, prevalence is interpreted as "the probability that the patient has the disease." This probability has to be assigned by the physician on the basis of clinical symptoms and/or prior testing.

This probabilistic interpretation is applicable to all the quantities in this unit that have been derived from the behavior of populations. The corresponding probabilities for the case of a single individual can be obtained by normalization (considering a total population Tot = 1).

2.4 Statistics of a Disease in Epidemiology

The prevalence of a disease in a population is related to two other quantities:

Incidence (I) of a disease, defined as number of cases of the disease per total population per year. The duration (Y) of a disease defined as the average time (in years) between the appearance of symptoms or identification by test, and complete remission or death (Figure 3.)

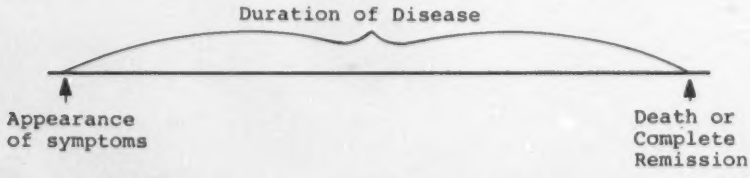


Figure 3. The duration of a disease is the average number of years between the appearance of symptoms, and complete remission or death.

The prevalence is then given by:

p(cases/population) = I(cases/(population × year)) × Y(years).

The incidence of a disease is assumed to be constant for a given population. On the other hand, the prevalence and duration of a disease may change if a sensitive clinical test is used to screen the population. The difference between times of detection by the test and by clinical symptoms is defined as lead time. Early detection by a test and early treatment may also change the duration of the disease (Figure 4).

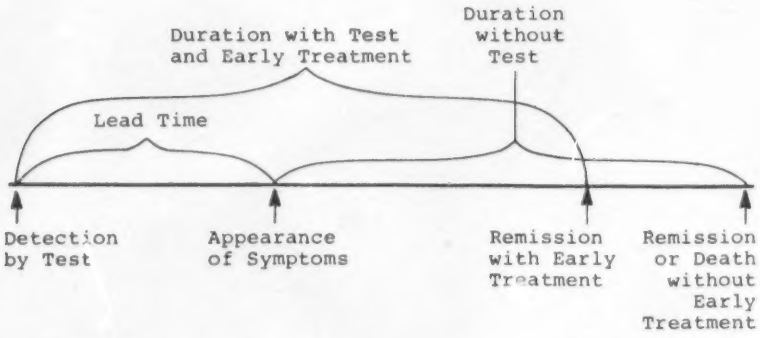


Figure 4. The time between the appearance of the symptoms of a disease and remission may be shortened by early detection and treatment.

A knowledge of the prevalence of a disease is essential in evaluating any clinical test designed to detect it in a large population. Thus, care must be taken to state clearly the conditions of test administration, and to pick the value of prevalence appropriate for these conditions. Factors to be considered in the assessment of a screening test include: distribution by age, sex, race and socioeconomic status of the population, and effect of lead time and early treatment.

2.5 A Procedure for Calculating Statistical Parameters

Given the size of the two populations (or probabilities) D and H, and the sensitivity (s) and specificity (c) of the test, each of the four populations (or probabilities) TP, TN, FP, and FN can be calculated using the definitions (1) through (4) and the relations (6) and (7):

(8) $TP = sD = spTot.$

(9) $TN = cH = c(1-p)Tot.$

(10) $FN = D - TP = (1-s)pTot.$

(11) $FP = H - TN = (1-c)(1-p)Tot.$

By definition, the total number of positive and negative results are given by:

(12) $Pos = TP + FP = [sp + (1-c)(1-p)]Tot.$

(13) $Neg = TN + FN = [c(1-p) + (1-s)p]Tot.$

While Equations (8) through (13) refer to populations, equations for the corresponding probabilities can be obtained by dividing by the total population (Tot). (Equivalently, Tot can be chosen to equal 1.)

The relationships above can be conveniently summarized in tabular form:

TABLE 1

Summary of Relationships

Population	Positive Results	Negative Results
Diseased D = pTot	TP = sD = spTot	FN = (1-s)D = (1-s)pTot
Healthy H = (1-p)Tot	FP = (1-c)H = (1-c)(1-p)Tot	TN = cH = c(1-p)Tot
Total Population Tot = H + D or Tot = Pos + Neg	Total Positive Pos = TP + FP = spTot +(1-c)(1-p)Tot	Total Negative Neg = TN + FN = c(1-p)Tot +(1-s)pTot

Note that a corresponding table of probabilities is obtained by setting Tot = 1.

The arrangement of Table 1 suggests a simple procedure for calculating TP, FN, FP, TN, Pos and Neg. given values of Tot, p, s and c:

a. Starting from the lower left-hand corner and using values of Tot and p appropriate to the situation, calculate D and H.
b. Using the value of D and the sensitivity s of the test, calculate TP and FN.
c. Similarly, using the value of H and the specificity c, calculate FP and TN.
d. Add up TP and FP to obtain Pos; similarly, add FN and TN to obtain Neg.

We shall illustrate this procedure by applying it to the following hypothetical situation:

Model Problem 1: A new test is to be used for screening a population for a disease that has a prevalence of 50/100,000. If the test is 80-percent sensitive and 90-percent specific, calculate the expected numbers of TP, FN, FP, and TN, Pos, and Neg when 100,000 subjects are screened.

We are given:

$$\begin{aligned} \text{Tot} &= 100{,}000 \\ p &= 50/100{,}000 \\ s &= 80\% \\ c &= 90\%. \end{aligned}$$

Using the procedure above we can construct the following table.

D = 50	TP = 0.8 × 50 TP = 40	FN = 0.2 × 50 FN = 10
H = 99,950	FP = 0.1 × 99,950 FP = 9,995	TN = 0.9 × 99,950 TN = 89,955
Tot = 100,000	Pos = 40 + 9,995 Pos = 10,035	Neg = 10 + 89,955 Neg = 89,965

2.6 Predictive Values

What the clinician is interested in is the reliability of a positive or negative finding. As a measure of this, two quantities are defined:

Predictive value of a positive result (PVP) = fraction of positive results that are true positives.

$$\text{PVP} = \frac{\text{TP}}{\text{Pos}} = \frac{sp}{sp + (1-c)(1-p)}. \tag{14}$$

Predictive value of a negative result (PVN) = fraction of negative results that are true negatives.

$$\text{PVN} = \frac{\text{TN}}{\text{Neg}} = \frac{c(1-p)}{c(1-p) + (1-s)p}. \tag{15}$$

These two equations are easily derived from (8) through (13).

Note that PVP and PVN are independent of the total population (Tot). Thus they are equally applicable to the epidemiological and individual situations as long as the appropriate value of p is used. PVP is the probability of disease given a positive test result, and can be used as

the measure of prevalence if further tests are being considered.

Model Problem 2: Calculate PVP and PVN for the situation described in Model Problem 1 (Section 2.4).

$$PVP = \frac{TP}{Pos} = \frac{40}{10,035} = 0.3986\%.$$

$$PVN = \frac{TN}{Neg} = \frac{89,955}{89,965} = 99.99\%.$$

These results imply that only 4 out of 1,000 positive results will be true positives, the remaining 996 being false positives. On the other hand, only 1 out of 10,000 negative results will be false negative, the remaining 9,999 being true negatives.

In the situation where the test is to be used for diagnosing an individual patient, the criterion of usefulness (Section 1.2) can now be expressed quantitatively. We require that the expected level of certainty of a positive result be greater than the present level of certainty of diagnosis; i.e.,

(16) $PVP > p.$

Alternatively, if the test result is to be used for elimination of a disease from consideration, we require that the predictive value of a negative result be greater than the present level of certainty for the absence of the disease; i.e.,

(17) $PVN > 1-p.$

These two conditions can be combined to give, i.e.,

(18) $PVP > p > 1-PVN.$

Model Problem 3: A physician is considering using a test that has 70-percent sensitivity and 90-percent specificity to confirm a diagnosis that she has made with 30-percent certainty on the basis of clinical symptoms of her patient. On the basis of the usefulness criterion alone, should she order this test for the patient?

We are given:

$s = 70\%$
$c = 90\%$
$p = 30\%.$

In this problem we shall use Equation (14) to calculate PVP.

$$PVP = \frac{sp}{sp + (1-c)(1-p)}$$

$$= \frac{0.70 \times 0.30}{0.70 \times 0.30 + 0.10 \times 0.70}$$

$$= \frac{0.21}{0.28} = 0.75\%.$$

Using (15),

$$PVN = \frac{c(1-p)}{c(1-p) + (1-s)p} = 87.5\%.$$

Therefore, the test satisfies the usefulness criterion (18) and would be justified on this basis alone (ignoring any costs or risks to the patient). If this physician had the choice between two different tests, each with a certain predictive value for a positive result, she should choose the test giving the higher PVP, unless she was actually interested in eliminating the possibility of a given disease. In the latter case she would start with an a priori probability of absence of disease, i.e., 1-p = 0.70, and should choose the test that provides her, under the conditions, a maximal PVN.

Equations (14) and (15) express the predictive values of positive and negative findings in terms of the parameters s, c, and p. By algebraic manipulations and examination of the equations, the following relationships can be established.

PVP increases unambiguously when s increases,
c increases,
or p increases.
PVN increases unambiguously when s increases,
c increases,
or p decreases.

3. MULTI-DISEASE AND MULTI-TEST ANALYSIS

3.1 Statistical Independence

In a real clinical situation, the diagnostic process may involve consideration of several diseases and the use of several clinical tests. The general situation can be analyzed in a manner similar to the analysis of a single test, provided the tests are all statistically independent. Two events are considered to be statistically independent if the probability of occurrence of one does not depend on whether or not the second event has occurred. If statistical independence holds, the probability of occurrence of both events is given by the product of the

probabilities of each. In other words, a set of tests is considered statistically independent if the probability of obtaining a positive (or negative) result on all of them is given by the product of the probabilities of obtaining positive (or negative) results on each one. This allows us to calculate the overall sensitivity and specificity of a battery of tests from the individual sensitivities and specificities. If the condition of statistical independence is not satisfied, the sensitivity and specificity of a battery of tests has to be estimated experimentally.

As an example, consider the use of three imaging techniques, namely conventional x-ray radiography, xerography and computerized tomography to establish the presence of a tumor in the lung. These are definitely not independent tests, since in each case we are measuring differences in absorption of x-rays. Adding ultrasonic imaging to the battery of tests would not introduce a statistically independent test either, since sonic reflection depends on differences in impedence which in turn depend on differences in density--a parameter which also affects x-ray absorption. Similarly, in the assessment of myocardial function, measurement of blood pressure and cardiac output would result in partially interdependent parameters. However, the shape of QRS complex in ECG and enzyme levels in blood may constitute independent measures of the state of the myocardium.

3.2 Series and Parallel Interpretation of Multiple Tests

When more than one test is to be used in the diagnostic process, a physician has a choice of interpreting a positive outcome on all the tests as a positive result (series interpretation) or interpreting a positive outcome on any one test as a positive result (parallel interpretation). The terms series and parallel are used in analogy with electrical circuits. If a set of switches are connected in series, the circuit is "on" only if all switches are "on." If the switches are connected in parallel, the circuit is "on" when any one of the switches is "on."

Let us consider two tests A and B. There are four possible outcomes of these two tests:

A^+B^+, A^+B^-, A^-B^+, A^-B^-.

In the series mode, A^+B^+ will be interpreted as a positive result while A^+B^-, A^-B^+, and A^-B^- will all be interpreted as negative.

In the parallel mode, A^+B^+, A^+B^-, and A^-B^+ will all be interpreted as positive results, and only A^-B^- will be interpreted as negative.

In the series interpretation of tests more false negatives will occur because a negative on any one test is interpreted as a negative result for the battery of tests. Under these circumstances, specificity will be high (fewer false positives) but sensitivity will be low (more false negatives). Conversely, in the parallel interpretation of tests sensitivity will be high (fewer false negatives) but specificity will be low (more false positives). Thus, the mode of interpretation of multiple tests should be carefully chosen to fit the diagnostic situation. High sensitivity is desired when the cost of false negatives is high, for instance, if the disease is dangerous and confirmation of positive results is relatively easy. On the other hand, if early diagnosis is not a significant advantage and false positives are excessively alarming and difficult to confirm, high specificity is desirable.

3.3 Notation

In this section we shall develop the necessary notation to describe a diagnostic situation where n mutually exclusive diseases are being considered, and m symptoms (some or all of which may be correlated to any of the diseases) can be monitored by means of m independent tests:

The prevalence (probability of occurrence) of disease d is denoted by

$$p_d; \quad d = 1,\ldots,n.$$

Assuming that the n diseases are mutually exclusive (i.e., a combination of different diseases in the same patient is not allowed), the total probability of disease is given by:

$$p(D) = \sum_{d}^{n} p_d$$

where the notation $\sum_{d}^{n}$ is used to indicate the sum of all terms from d = 1 to d = n.

The probability of nondisease (i.e., being healthy) is then given by:

$$p(H) = 1 - p(D) = 1 - \sum_{d}^{n} p_d.$$

The sensitivity of test t for disease d (probability of existence of symptom t given the presence of disease d) is denoted by

$$s_{td}, \quad t = 1,\ldots,m; \quad d = 1,\ldots,n.$$

The specificity of test t (probability of absence of symptom t given the absence of all n diseases under consideration) is denoted by

$$c_t, \quad t = 1,\ldots,m.$$

Note that the specificities c_t refer to the population that is free of <u>all</u> the diseases being considered, and is, therefore, a measure of the absence of the symptom in the general, healthy population. This approach is valid as long as all relevant disease states are included in the list of n diseases.

The <u>overall</u> sensitivity of the m tests for disease d (denoted by S_d) and their overall specificity (denoted by C), can be calculated from the individual sensitivities s_{td} and specificities c_t.

3.4 Calculation of Overall Sensitivity and Specificity

Assuming that the m tests are statistically independent, the overall sensitivity and specificity of the battery of tests can be calculated for the series and parallel interpretation of the tests.

<u>Series Interpretation</u>: The result of the battery of tests is considered positive only if all tests yield a positive outcome. Thus, the probability of obtaining a true positive result on the battery of tests will be given by the product of probabilities of obtaining a true positive result on each test:

$$(19) \qquad S_d = \prod_t^m s_{td}.$$

The notation $\prod_t^m s_{td}$ is used to indicate the product $s_{1d}s_{2d}s_{3d}\cdots s_{md}$.

Similarly, the probability of obtaining a false positive on the battery of tests will be given by the product of probabilities of obtaining a false positive on each test:

$$(20) \qquad 1 - C = \prod_t^m (1-c_t)$$

and

$$(21) \qquad C = 1 - \prod_{t}^{m}(1-c_t).$$

Parallel Interpretation: In this case the result of the battery of tests is considered negative only if all tests yield negative outcomes. The overall probability of a false negative or a true negative can, therefore, be expressed as products of individual probabilities:

$$(22) \text{ False negatives: } \quad 1 - S_d = \prod_{t}^{m}(1-s_{td}).$$

$$(23) \text{ Therefore: } \quad S_d = 1 - \prod_{t}^{m}(1-s_{td}).$$

$$(24) \text{ True negatives: } \quad C = \prod_{t}^{m} c_t.$$

Again note that the overall sensitivities S_d refer to the individual diseases, while the overall specificity C refers to the healthy population (i.e., free of all the diseases).

If a single disease is being considered in conjunction with multiple tests, the overall sensitivities and specificity calculated for either the concurrent or sequential modes can be used directly in Equations (14) and (15) to calculate predictive values. The multi-disease situation is analyzed in detail in Section 3.5.

Model Problem 4: We shall illustrate the application of this formalism by considering the hypothetical case of a physician suspecting two alternative diseases. Three symptoms can be monitored by means of three clinical tests. Given the following sensitivities and specificities:

Sensitivities for:

	Disease 1	Disease 2	Specificities
Test 1:	$s_{11} = 0.90$	$s_{12} = 0.50$	$c_1 = 0.75$
Test 2:	$s_{21} = 0.80$	$s_{22} = 0.10$	$c_2 = 0.90$
Test 3:	$s_{31} = 0.60$	$s_{32} = 0.85$	$c_3 = 0.50$

Calculate the overall sensitivities (for each disease) and overall specificities of the tests when interpreted in series and in parallel.

Series Interpretation: The overall sensitivity of the tests for the two diseases are given by Equation (19):

$$S_1 = \prod_t^3 s_{t1} = 0.90 \times 0.80 \times 0.60 = 0.432 \text{ or } 43.2\%.$$

$$S_2 = \prod_t^3 s_{t2} = 0.50 \times 0.10 \times 0.85 = 0.0425 \text{ or } 4.25\%.$$

Similarly, the overall specificity can be calculated from Equation (20):

$$1 - C = \prod_t^3 (1-c_t)$$

$$= (1-0.75)(1-0.90)(1-0.50)$$

$$= 0.0125$$

$$C = 1 - 0.0125 = 0.9875 \text{ or } 98.75\%.$$

Parallel Interpretation: Overall sensitivities are given by Equation (22):

$$1 - S_1 = \prod_t^3 (1-s_{t1})$$

$$= (1-0.90)(1-0.80)(1-0.60)$$

$$= 0.008$$

$$S_1 = 0.992 \text{ or } 99.2\%.$$

$$1 - S_2 = \prod_t^3 (1-s_{t2})$$

$$= (1-0.5)(1-0.1)(1-0.85)$$

$$= 0.0675$$

$$S_2 = 0.9325 \text{ or } 93.25\%.$$

Overall specificity is given by Equation (24):

$$C_1 = \prod_t^3 c_{t1}$$

$$= 0.75 \times 0.90 \times 0.50$$

$$= 0.3375 \text{ or } 33.75\%.$$

These results illustrate the fact that series interpretation of tests yields high specificity and low sensitivity, while parallel interpretation yields high sensitivity and low specificity.

3.5 Predictive Values for Multi-Disease Situation

In deriving predictive values of a test result (or set of results) for a particular disease, the contribution of false results of all other states (other diseases or health) have to be accounted for. Thus, the predictive value of a positive for disease d (i.e., the probability of disease d being present given a positive result) is given by:

$$PVP_d = \frac{\text{True positives for disease d}}{\text{All positive results}}$$

$$= \frac{\text{True positives for disease d}}{\text{Positives for all diseases + false positives}}$$

$$(25)\quad PVP_d = \frac{p_d S_d}{\sum_d^n p_d S_d + (1-\sum_d^n p_d)(1-C)},$$

where

S_d is the sensitivity of a single test or battery of tests for disease d.

$p_d S_d$ is the fraction of positives that are true positives for disease d.

$\sum_d^n p_d S_d$ is the fraction of positives contributed by all the disease.

$1-\sum_d^n p_d$ is the probability of having none of the diseases.

$(1-\sum_d^n p_d)(1-C)$ is the probability of a positive result when none of the diseases are present.

It may also be useful to know the probability of any of the n diseases being present, given a positive result. This probability is the predictive value of a positive result for the presence of any one of the n diseases:

$$(26)\quad PVP(\text{any d}) = \frac{\sum_d^n p_d S_d}{\sum_d^n p_d S_d + (1-\sum_d^n p_d)(1-C)}.$$

In a similar manner, the predictive value of a negative result may refer to the absence of a particular disease d, or the absence of all disease, given a negative

TABLE 2

Summary of Relationships for Multi-Disease System

State of Patient	Positive Results	Negative Results
Disease 1 p_1	True Positives of Disease 1 $TP_1 = p_1 S_1$	False Negatives of Disease 1 $FN_1 = p_1(1-S_1)$
Disease 2 p_2	True Positives of Disease 2 $TP_2 = p_2 S_2$	False Negatives of Disease 2 $FN_2 = p_2(1-S_2)$
. . .	. . .	. . .
Disease d p_d	True Positives of Disease d $TP_d = p_d S_d$	False Negatives of Disease d $FN_d = p_d(1-S_d)$
. . .	. . .	. . .
Disease n p_n	True Positives of Disease n $TP_n = p_n S_n$	False Negatives of Disease n $FN_n = p_n(1-S_n)$
Healthy $p(H) = 1-\sum_d^n p_d$	False Positives of Healthy $FP = (1-\sum_d^n p_d)(1-C)$	True Negatives of Healthy $TN = (1-\sum_d^n p_n)C$
Total $p(H) + \sum_d^n p_d = 1$	Total Positives $\sum_d^n p_d S_d + (1-\sum_d^n p_d)(1-C)$	Total Negatives $\sum_d^n p_d(1-S_d) + (1-\sum_d^n p_n)C$

result. The predictive value of a negative result for the absence of all n diseases is given by:

$$(27)\quad PVN\ (\text{all } d) = \frac{(1-\sum_{d}^{n} p_d)C}{(1-\sum_{d}^{n} p_d)C + \sum_{d}^{n} p_d(1-S_d)}.$$

The predictive value of a negative result for the absence of a particular disease d is given by:

PVN_d = fraction of all negatives that are true negatives for disease d

= 1 - fraction of all negatives that are false negatives for disease d

$$(28)\quad PVN_d = 1 - \frac{p_d(1-S_d)}{(1-\sum_{d}^{n} p_d)C + \sum_{d}^{n} p_d(1-S_d)}.$$

These relationships may be summarized in tabular form (see Table 2).

Model Problem 5: Referring to model problem 4, and assuming that the physician assigned an a priori probability of 10 percent for disease 1 to be present, and 20 percent for disease 2 to be present, calculate the following predictive values for concurrent administration of the three tests:

PVP_1; PVP_2; PVP(1 or 2); PVN_1; PVN_2; PVP(1 and 2).

Let us recall that for concurrent administration,

$S_1 = 43.2\%$; $S_2 = 4.25\%$; $C = 98.75\%$.

The tabular representation of Model Problem 5 is as follows:

State of Patient	Positive Results	Negative Results
Disease 1 $P_1 = 0.10$	$TP_1 = P_1S_1$ $= 0.10\times0.432$ $= 0.0432$	$FN_1 = P_1(1-S_1)$ $= 0.10\times0.568$ $= 0.0568$
Disease 2 $P_2 = 0.20$	$TP_2 = P_2S_2$ $= 0.20\times0.0425$ $= 0.0085$	$FN_2 = P_2(1-S_2)$ $= 0.20\times0.9575$ $= 0.1915$
Healthy $p(H) = 1-(0.10+0.20)$ $= 0.70$	$FP = p(H)(1-C)$ $= 0.70\times0.0125$ $= 0.00875$	$TN = p(H)C$ $= 0.70\times0.9875$ $= 0.69125$
Total 1	Total Positives 0.06045	Total Negatives 0.93955

Corresponding to Equation (25),

$$PVP_1 = 0.0432/0.06045 = 0.7146$$
$$PVP_2 = 0.0085/0.06045 = 0.1416$$

Equation (26),

$$PVP(1 \text{ or } 2) = (0.0432 + 0.0085)/0.06045 = 0.8553$$

Equation (28),

$$PVN_1 = 1 - (0.0568/0.93955) = 0.9396$$
$$PVN_2 = 1 - (0.1915/0.93955) = 0.7962$$

Equation (27),

$$PVN(1 \text{ and } 2) = 0.69125/0.93955 = 0.7357.$$

3.6 Predictive Value of an Arbitrary Set of Tests Results

A more general situation that can be described by the probabilistic approach consists of the following:

Let $D_1 \ldots D_n$ be n mutually exclusive disease states that are being considered in the given diagnostic situation.

Let $T_1 \ldots T_m$ be m independent tests relating to one or more of these diseases.

Then, s_{td} is the sensitivity of test t for disease d, and c_t is the specificity of test t for absence of all n diseases.

Series or parallel interpretations of the m tests would be useful if all the tests were related to a given disease. In the modern practice of obtaining biochemical multi-test profiles by automated systems, only a few of the many measured parameters are known to be related to a given disease. Under these conditions it is desirable to know the predictive value of any <u>set</u> of test results for any one of the diseases. To achieve this we need to define a new variable a_t, $t = 1,\ldots,m$, such that $a_t = 1$ when test t is positive, and $a_t = 0$ when test t is negative. Then the predictive value of a given set of test results for disease d is given by:

$$PV_d = \frac{P_d \prod_t [a_t s_{td} + (1-a_t)(1-s_{td})]}{\sum_i P_i \prod_t [a_t s_{ti} + (1-a_t)(1-s_{ti})] + (1-\sum_i P_i) \prod_t [a_t(1-c_t) + (1-a_t)c_t]}$$

where $\prod_t$ indicates the product of all terms as t is varied from 1 to m, and $\sum_i$ indicates the sum of all terms as i is varied from 1 to n. Note that this predictive value does

not refer to a positive or negative result but instead expresses the probability of presence of disease d for any combination of positive and negative results on the individual tests.

The variable a_t was used here as an "on-off switch" to indicate the presence or absence of symptom t according to the test. This same variable a_t is useful in another context: a value of a_t between 0 and 1 can be used to express the level of certainty associated with the presence of a symptom. For instance, examination of a patient's EKG may be best represented by a "probability that the EKG is abnormal" of 70 percent. The corresponding a_t can then be set at 0.70.

4. MEDICAL DECISION ANALYSIS

4.1 Medical Decisions and Strategies

Medical decisions involve choices of diagnostic and/or therapeutic procedures. A strategy consists of a sequence of such choices, taking into account all possible outcomes. A simple example of a strategy is: a patient is suspected to have strep infection of the throat. The physician decides to perform a strep culture test, planning to indicate treatment with an antibiotic if the test result is positive, or no treatment at all if it is negative. The two decisions (choice of test and choice of treatment) are connected by two probabilistic events: probability of presence or absence of disease, and probability of obtaining a positive or negative result. The decision to treat is followed by the probability of side effects while the deci-

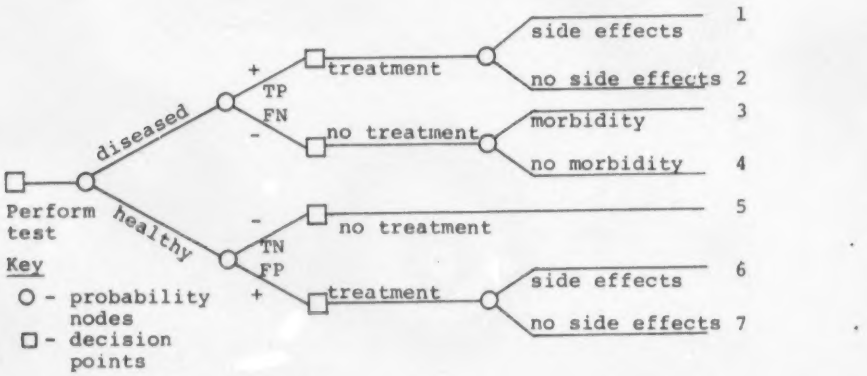
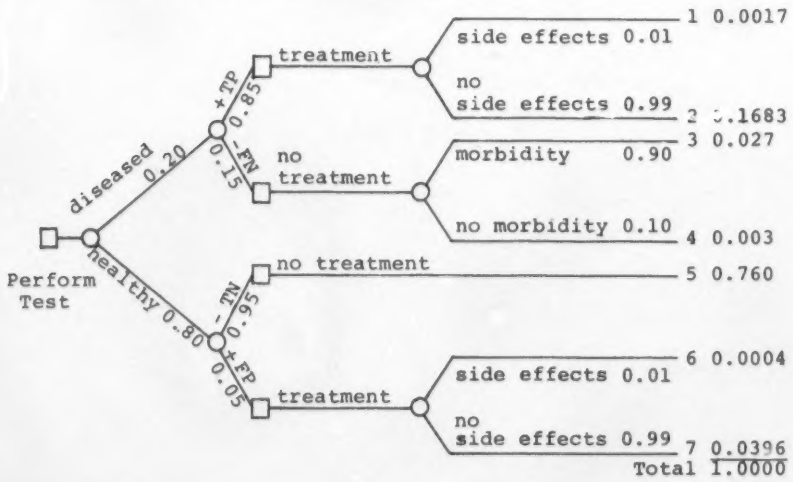


Figure 5. Decision tree when a clinical test is indicated, taking into account the possibility of side effects from treatment.

sion not to treat a false negative is followed by probability of morbidity occurring. A decision tree represents decisions and the connecting probabilistic events as decision points and probability nodes in a line diagram (Figure 5).

Note that the strategy represented by Figure 5 yields seven different outcomes. The probability of each outcome is given by the product of probabilities of all the branches leading to that outcome.

Model Problem 6: Given that the physician assigned a prevalence of 20 percent to the disease on the basis of an examination, the sensitivity of the test is 85 percent, the specificity of the test is 95 percent, the probability of side effects from the treatment is 1 percent, and probability of morbidity resulting when the disease is left untreated is 90 percent, assign probabilities to each branch of probability nodes in Figure 5, and calculate the overall probabilities of the seven possible outcomes. These probabilities are shown in the figure below.



It is important to point out that the number and nature of decision points in a given strategy depend on the alternatives that are included in the consideration of the problem. While it is clear that irrelevant or obviously unacceptable alternatives should not be included, the physician should make sure that all relevant and reasonable alternatives are taken into account in the analysis of the problem. For instance, the decision tree in Figure 5 excludes the alternatives of treatment for negative test results and nontreatment for positive test results. This

exclusion may or may not be justified, depending on the severity of the disease, the cost and risks of treatment, and the expected certainty of test results.

4.2 Costs, Benefits, and Weighting Coefficients

To evaluate a particular strategy, costs and benefits related to each of the diagnostic or therapeutic procedures involved have to be taken into account. Below is a list of factors that may be included in a cost/benefit analysis:

Costs

Financial costs (e.g., cost of tests and treatments)
Psychological stress (e.g., due to false test results or dangerous treatments)
Physical danger (e.g., exposure to radiation, chemotherapy)
Decrease in life expectancy (e.g., removal of a critical organ)
Increased chance of death (e.g. exploratory surgery)
Social stress (e.g., loss of job)

Benefits

Increase in life expectancy (e.g. successfully treated cancer)
Reduced chance of death (e.g., use of cardiac pacemaker)
Reduced pain and suffering (e.g. radiation treatment)
Reduced level of disability (e.g., use of supportive therapy)
Elimination of psychological stress (e.g., testing for a suspected lethal disease).

The list above is not meant to be exhaustive, mainly because the consideration of costs and benefits have to reflect the concerns and values of the individual patient.

Another instance where individualized tailoring is necessary is in the choice of weighting coefficients. While each of the cost and benefit factors may be quantitated, the medical specialist has to assign weighting coefficients to each item so that all costs and benefits can be expressed in the same units and an overall cost/benefit function can be obtained. The common unit may be dollars, life expectancy, fatality rate, level of disability, or an arbitrary unit. Whatever the choice of units, the normalization procedure is definitely an over-simplification of the complex issues involved in a medical decision. However, the normalization is necessary for a quantitative comparison of alternative strategies.

The normalized costs and benefits of the steps leading to the ith outcome can be added up to yield the overall utility U(i) of the ith outcome. In this procedure, benefits will appear as positive quantities, while costs will appear as negative quantities.

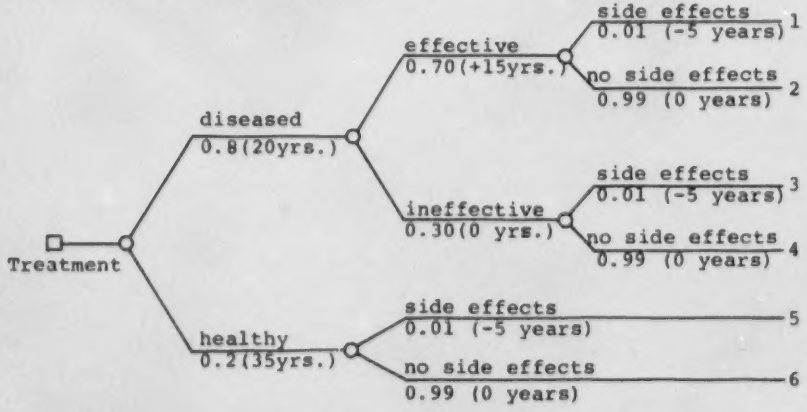
Then, the overall cost/benefit F for the strategy is given by:

$$F = \sum_{i}^{N} q_i U(i), \tag{31}$$

where N is the total number of different outcomes of the strategy, and q_i is the probability of the ith outcome.

Thus, decision analysis can assist the physician in evaluating all the reasonable strategies in a given situation, particularly when risk factors are high and the choices not immediately clear.

Model Problem 7: Consider the case of a patient that has been diagnosed to have a disease with 80 percent certainty. The physician decides to use a treatment which has a 70 percent probability of success. If the treatment is ineffective, life expectancy will be 20 years; if effective, life expectancy will be the same as for a healthy subject (35 years). By undergoing this treatment, the patient runs a 1-percent risk of developing side effects that would reduce life expectancy by 5 years. Calculate the overall cost/benefit F for this strategy (decision to treat). In this example the normalized cost/benefit parameter is life expectancy.



The decision tree above shows all the probability nodes, and the probability and change in life expectancy of each branch. We can calculate the probability and utility (life expectancy) of each of the six outcomes:

$q_1 = 0.8 \times 0.7 \times 0.01 = 0.0056$ $\quad U(1) = 30$ yrs.

$q_2 = 0.8 \times 0.7 \times 0.99 = 0.5544$ $\quad U(2) = 35$ yrs.

$q_3 = 0.8 \times 0.3 \times 0.01 = 0.0024$ $\quad U(3) = 15$ yrs.

$q_4 = 0.8 \times 0.3 \times 0.99 = 0.2376$ $\quad U(4) = 20$ yrs.

$q_5 = 0.2 \times 0.01 = 0.002$ $\quad U(5) = 30$ yrs.

$q_6 = 0.2 \times 0.99 = 0.198$ $\quad U(6) = 35$ yrs.

The overall cost/benefit function of this strategy is:

$$F = 0.0056(30) + 0.5544(35) + 0.0024(15) + 0.2376(20) + 0.002(30) + 0.198(35)$$

$F = 31.35$ years (life expectancy).

4.3 Cost/Benefit Analysis of Treatment: Prevalence Threshold

Model Problem 7 refers to a common and important medical decision situation: given that a diagnosis has been made with probability (p), and the costs and benefits of the treatment under consideration are known, is the treatment indicated or counterindicated?

This situation can be represented by the two simple decision trees in Figure 6.

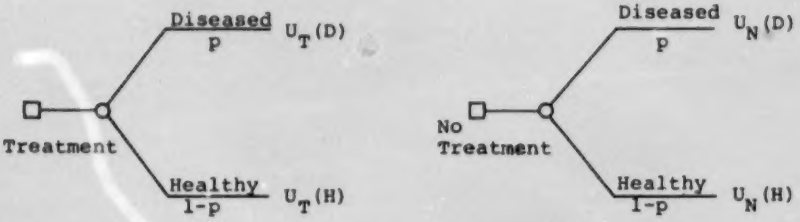


Figure 6. Decision trees for analysis of treatment.

The quantities $U_T(D)$, $U_T(H)$ are defined as the utility (normalized) of treatment if the subject is diseased and healthy, respectively. Similarly, $U_N(D)$ and $U_N(H)$ are the utilities of no treatment if the subject is diseased or healthy. Then, the overall cost/benefit function for treatment F_T is given by:

$$F_T = pU_T(D) + (1-p)U_T(H) \tag{32}$$

and the corresponding function F_N for no treatment is given by:

(33) $$F_N = pU_N(D) + (1-p)U_N(H).$$

Treatment would be indicated if $F_T > F_N$, and treatment would be counterindicated if $F_T < F_N$. When $F_T = F_N$, neither strategy is favored. At this point of indifference, we have the condition

(34) $$pU_T(D) + (1-p)U_T(H) = pU_N(D) + (1-p)U_N(H).$$

Solving for p:

(35) $$p = \frac{U_N(H) - U_T(H)}{U_T(D) - U_N(D) + U_N(H) - U_T(H)}.$$

Defining

(36) $$A = \frac{U_T(D) - U_N(D)}{U_N(H) - U_T(H)},$$

we obtain:

(37) $$p = \frac{1}{A + 1}.$$

The quantity (A) can be interpreted as the ratio of the advantage of treatment when the subject is diseased ($U_T(D) - U_N(D)$) over the advantage of no treatment when the subject is healthy ($U_N(H) - U_T(H)$). For this reason, A is called the advantage factor of treatment over no treatment. The value of (p) defined by Equation (37) is called the threshold value of p. It is interpreted as the minimum value of prevalence necessary to justify treatment. If $p > 1/(A+1)$, then $F_T > F_N$ and treatment is indicated. If $p < 1/(A+1)$, then $F_T < F_N$ and treatment is counterindicated. Figure 7 is a plot of threshold value p, vs. A.

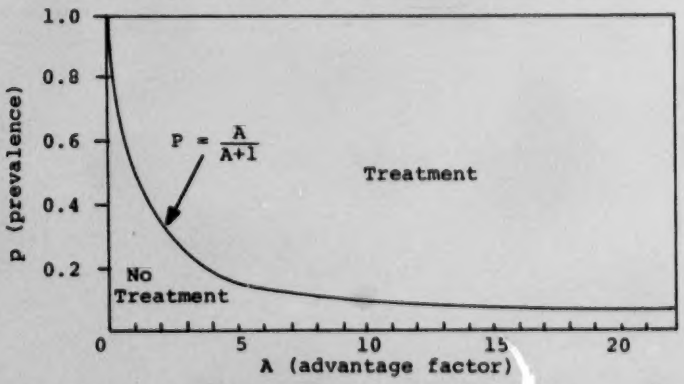


Figure 7. Plot of the threshold probability $p = \frac{A}{A+1}$, versus the advantage A.

This reasoning implies that treatment should be deferred until p exceeds the threshold value. Diagnostic tests have to be performed, therefore, until this level of certainty (indicated by PVP) is attained. Unless an alternative treatment which requires a higher certainty of diagnosis is being considered, further tests should not be performed since they would not change the strategy but will add to the overall cost.

5. EXERCISES

1. Referring to Figure 2, discuss the expected effect, on the following quantities, of moving the demarcation line towards the mean of the diseased population:

   TN FP TP FN
   Pos Neg D H
   PVP PVN

   What changes are expected if the demarcation line is moved towards the mean of the healthy population?

2. A newly developed test was administered to a group of 500 subjects diagnosed to have the disease, and another group of 800 healthy, asymptomatic subjects. The test gave 470 positive and 30 negative results in the first group, and 80 positive and 720 negative results in the second group.
   a. Calculate the sensitivity and specificity of the test.
   b. What additional information would you need to evaluate the usefulness of this test for an individual patient?
   c. What additional information would you need to evaluate the usefulness of this test as a screening test?

3. A patient has been diagnosed (with 5-percent certainty) to have syphilis. The Wasserman Test for syphilis has a sensitivity of 95 percent and specificity of 50 percent.
   a. Can the test be useful for confirming this diagnosis?
   b. Can the test be useful for eliminating syphilis from consideration?

4. After an x-ray examination, a patient was diagnosed to have tuberculosis with 75-percent certainty. If the examination is 90-percent sensitive and 70-percent specific, what was the level of certainty of diagnosis prior to the test?

5. A diagnosis has been made with 40-percent certainty on the basis of clinical symptoms. What sensitivity and specificity is required from a test to give a PVP and PVN of 80 percent?

6. The prevalence of cancer of the colon is 12/100,000 in adult males, the average duration of the disease is 4 months (from time of manifestation of clinical symptom to the terminal stage or to virtually complete remission). A test has been devised which detects this cancer earlier. Examining a representative population the prevalence of the disease was found to be 18/100,000. What is the average lead time for early detection of this disease (in months)?

7. Calculate the proportion of false results when a test with 95-percent sensitivity and 95-percent specificity is used and the prevalence is 10 percent. What will be the proportion of false

results if three tests, each with sensitivity 95 percent and specificity 95 percent, are used in parallel? Repeat the calculation for series use. What does this imply about the use of multiple tests for diagnosis?

8. Referring to the information given in Model Problems 4 and 5, calculate PVP_1, PVP_2, PVP (1 or 2), PVN_1, PVN_2, and PVN (1 and 2) under the following conditions.
   a. Tests 1 and 2 (series and parallel);
   b. Tests 1 and 3 (series and parallel);
   c. Tests 2 and 3 (series and parallel).

   Which choice and mode of administration would suit these purposes:
   d. Optimizing PVP for disease 1;
   e. Optimizing PVP for disease 2;
   f. Optimizing PVP for disease 1 or 2;
   g. Optimizing PVN for disease 1;
   h. Optimizing PVN for disease 2;
   i. Optimizing PVN for disease 1 and 2.

9. In a pilot study on women, ages 35-45, mammography and physical examination (palpation) were used for early detection of incipient breast cancer, which has a prevalence of 1 percent in this population. Physical examination alone detected 52 percent of confirmed cases, while mammography alone detected 65 percent of the cases. When the two techniques were used in series, 42 percent of cancers were detected; when used in parallel, 75 percent of cancers were detected.

   Confirmation of breast cancer was done by biopsy. Only 1 in 5 women sent to biopsy on the basis of physical examination alone, and 1 in 4 of those indicated by mammography alone, had breast cancer. The biopsies confirmed 25 percent of those diagnosed by series use of the tests, and 18 percent of those diagnosed by parallel use of the tests.
   a. Are these two tests statistically independent?
   b. Calculate the specificities of: mammography, physical examination, the two used in series, and the two used in parallel.
   c. Assuming a cost of $20 per physical examination, $40 per mammography, and $400 per biopsy, what is the cost per cancer detected for each of these five situations: physical alone, mammography alone, physical and mammography in series, physical followed by mammography, mammography followed by physical. (Note: <u>all</u> positive results have to be confirmed by biopsy.)
   d. Repeat the analysis in (c) assuming an added cost of $500 for the risk of radiation-induced cancer by mammography, and $10,000 for each case missed by the diagnostic procedure.

10. For Model Problem 7, draw a decision tree and calculate F for the strategy of leaving the patient untreated. Is the physician's decision to treat the patient justified?

11. Calculate the advantage factor (A) and the threshold probability (p) for the situation of Model Problem 7. With the 80-pecent certainty of diagnosis, is treatment indicated?

12. For the strategy described by Figure 5 and Model Problem 6, calculate the overall cost/benefit function using disability as the normalized cost. (Assume treated disease has 5-percent disability, untreated disease with morbidity 90-percent disability, untreated disease without morbidity 10-percent disability, side effects 15-percent disability.)

13. An ordinary sore throat may be considered as caused either by bacterial or viral infection. For this problem, assume that sore throat is caused either by streptococcus (25 percent of cases) or by a virus (75 percent of cases), but not both. The treatment of a viral infection is purely symptomatic; i.e., bedrest and gargling. A strep infection has complications that result in 0.1-percent fatality if the infection runs its course. Treatment of strep with chloramphenicol is 95-percent effective, but 1/30,000 subjects exposed to the antibiotic develop a fatal aplastic anemia. If the antibiotic is given from the first day, the probability of complications due to the strep drops 15-fold. If given from the second day onward, the probability of complications drops to 1/2 of that of the untreated infection.

    A throat culture, that requires one day to analyze, can be used to detect the strep throat. The test is 75-percent sensitive and 90-percent specific. As a physician treating a patient with sore throat, consider the following questions:

    a. What are the alternative strategies you can adopt?
    b. What is the most important cost/benefit factor you should take into account?
    c. For each strategy in (a), draw a decision tree and calculate the overall cost/benefit function. Which strategy is the preferred one?

6. SAMPLE EXAM

1. Describe the experimental procedure for determining the sensitivity and specificity of a diagnostic test. Explain the importance of choice of population and the demarcation line in this procedure.

2. Define prevalence of a disease. How is it determined for a given population, and for an individual patient? Will the introduction of lead time by a new test make any change in incidence or prevalence of the disease?

3. Compare predictive values of positive and negative results to the concepts of sensitivity and specificity. Which of these quantities depend on prevalence and how?

4. How would you experimentally determine PVP and PVN of a diagnostic procedure?

5. What is the difference between decision points and probability nodes?

6. What are the subjective elements involved in constructing decision trees and performing cost/benefit analyses? Are these elements introduced by the physician or the patient?

7. Explain the concepts of "advantage factor of treatment over nontreatment" and "prevalence threshold." How are they related?

8. A new serological screening test for gonorrhea has been proposed by a company. The test has a sensitivity of 80 percent and specificity of 95 percent. Calculate the predictive value of a positive and a negative result on the test for:
   a. Screening asymptomatic women in the 20-24 year age group (prevalence = 2%).
   b. Screening asymptomatic women in the 25-29 age group (prevalence = 1%).
   c. Diagnosing a patient with 10-percent probability of having gonorrhea on the basis of symptoms.

   In which situation would you consider the test to be most useful?

9. The prevalence of renovascular disease in hypertensive patients is 10 percent. What sensitivity and specificity are required of intravenous pyelography (IVP) if a PVP of 80 percent and a PVN of 99.9 percent is desired? If IVP has a sensitivity of 99 percent and specificity of 90 percent in detection of renovascular hypertension, does it satisfy the above requirements?

10. Measurement of blood pressure (BP), history taking (H), and electrocardiography (ECG) can all be used for the detection of heart disease. The sensitivities and specificities of the three

tests are tabulated below:

	S	C
BP	85%	75%
H	55%	60%
ECG	60%	80%

In what mode (series or parallel) would you combine these three tests for the purpose of screening a large population? Calculate the expected combined s and c, for both series and parallel modes, assuming the three procedures are statistically independent. What would be the expected PVN, for both modes, if the tests are used for screening a population with 5-percent prevalence of heart disease?

11. On the basis of preliminary tests, a patient is suspected (with 5-percent probability) to have a malfunctioning mitral valve. If left untreated, the condition results in early mortality in 30 percent of cases. The test that can confirm this suspicion is 98-percent sensitive, 95-percent specific, but carries with it 1-percent chance of mortality. Surgical treatment has 5-percent chance of mortality during surgery, but is otherwise a reliable procedure. Compare the following two strategies by drawing decision trees and calculating the expected mortality:
    a. No further testing or treatment is undertaken.
    b. Perform the confirmatory test and treat surgically if the result is positive.

    Calculate the advantage factor of treatment over nontreatment and the threshold probability for this situation.

7. ANSWERS TO EXERCISES

1. D and H are independent of demarcation line and will show no change. Neg, TN, FN, and PVP will increase if line is moved towards the mean of the diseased population, while Pos, TP, FP, and PVN will decrease. Opposite changes are expected when demarcation line is moved towards the mean of the healthy population.

2. s = TP/D = 470/500 = 94%
   c = TN/H = 720/800 = 90%.

3. To evaluate usefulness of the test, PVP and PVN should be calculated. We can use Equations (14) and (15), with p = 5%, s = 95%, c = 50%, or the table below.

D = 5	TP = 0.95 × 5 = 4.75	FN = (1-0.95)5 = 0.25	s = 95%
H = 95	FP = (1-0.5)95 = 47.5	TN = 0.5 × 95 = 47.5	c = 50%
Tot = 100	Pos = 52.25	Neg = 47.75	
P = 5%	PVP = 4.75/52.25 = 9.1%	PVN = 47.5/47.75 = 99.5%	

   The criterion of usefulness (18) is:

$$PVP > p > 1 - PVN.$$

   Since .091 > .05 > 1 - .995, the test is useful for both confirming or eliminating the disease.

4. s = 90%, c = 70%, PVP = 75%, p = ?

   Using Equation (14): $0.75 = \dfrac{0.9p}{0.9p + 0.3(1 - p)}$.

   Rearranging and solving for p: p = 0.5 = 50%.

5. p = 40%, PVP = PVN = 80%, s = ?, c = ?

   Using (14) and (15):

$$0.8 = \frac{0.4s}{0.4s + 0.6(1 - c)} \quad \text{and} \quad 0.8 = \frac{0.6c}{0.6c + 0.4(1 - s)}.$$

$$8s + 12(1 - c) = s \quad \text{and} \quad 24c + 16(1 - s) = 3c$$

$$7s - 12c = -12 \qquad 21c - 16s = -16.$$

   Solving the two equations simultaneously,

$$s = 66.7\% \quad \text{and} \quad c = 88.9\%.$$

6. p = 12/100,000 when D = 4 mo.

$$I = p/D = 3/100{,}00 \text{ per month}$$

   With new test, p = 18/100,000

new duration D = p/I = 6 mo.

Lead time = 6 - 4 = 2 mo.

7. p = 10%, s = 95%, c = 95%

D = 10	TP = 9.5	FN = 0.5	s = 95%
H = 90	FP = 4.5	TN = 85.5	c = 95%
Tot = 100	Pos = 14	Neg = 86	

% false results = FP + FN = 5%

Three tests in series: $S = (0.95)^3 = 85.74\%$

$C = 1 - (0.05)^3 = 99.99\%$

D = 10	TP = 8.574	FN = 1.426	S = 85.74%
H = 90	FP = 0.011	TN = 89.989	C = 99.99%
Tot = 100	Pos = 8.585	Neg = 91.415	

% false results = 1.437%

Three tests in parallel: S = 99.99%

C = 85.74%

D = 10	TP = 9.999	FN = 0.001	S = 99.99%
H = 90	FP = 12.834	TN = 77.166	C = 85.74%
Tot = 100	Pos = 22.833	Neg = 77.167	

% false results = 12.835%

8. $S_{11} = 0.90$ $S_{12} = 0.50$ $C_1 = 0.75$ $P_1 = 10\%$
$S_{21} = 0.80$ $S_{22} = 0.10$ $C_2 = 0.90$ $P_2 = 20\%$
$S_{31} = 0.60$ $S_{32} = 0.85$ $C_3 = 0.50$ p(H) = 70%

Combined sensitivities and specificities using (19)-(24):

	In Series			In Parallel		
Tests	S_1	S_2	C	S_1	S_2	C
1 and 2	0.72	0.05	0.975	0.98	0.55	0.675
1 and 3	0.54	0.425	0.875	0.96	0.925	0.375
2 and 3	0.48	0.875	0.95	0.92	0.865	0.45

a. Tests 1 and 2

Series

	+	-	
D_1 = 10	1.2	2.8	S_1 = 0.72
D_2 = 20	1.0	19.0	S_2 = 0.05
H = 70	1.75	68.25	C = 0.975
100	9.95	90.05	

Parallel

	+	-	
D_1 = 10	9.8	0.2	S_1 = 0.98
D_2 = 20	11.0	9.0	S_2 = 0.55
H = 70	22.75	47.25	C = 0.675
100	43.55	56.45	

$PVP_1 = 0.724$ $PVN_1 = 0.969$

$PVP_2 = 0.100$ $PVN_2 = 0.789$

PVP = .824 PVN = 0.765

(1 or 2) (1 and 2)

$PVP_1 = 0.225$ $PVN_1 = 0.996$

$PVP_2 = 0.253$ $PVN_2 = 0.841$

PVP = 0.478 PVN = 0.837

(1 or 2) (1 and 2)

b. and c. Use same procedure, but with values of S and C for 1 and 3, and 2 and 3, respectively.

b. 1 and 3

Series		Parallel	
PVP_1 = 0.238	PVN_1 = 0.940	PVP_1 = 0.134	PVN_1 = 0.986
PVP_2 = 0.375	PVN_2 = 0.851	PVP_2 = 0.259	PVN_2 = 0.947
PVP = 0.613 (1 or 2)	PVN = 0.792 (1 and 2)	PVP = 0.392 (1 or 2)	PVN = .932 (1 and 2)

c. 2 and 3

Series		Parallel	
PVP_1 = 0.48	PVN = 0.942	PVP_1 = 0.142	PVN_1 = 0.977
PVP_2 = 0.17	PVN = 0.797	PVP_2 = 0.266	PVN_2 = 0.923
PVP = 0.65 (1 or 2)	PVN = 0.739 (1 and 2)	PVP = 0.408 (1 or 2)	PVN = .900 (1 and 2)

d. 1 and 2 series
e. 1 and 3 series
f. 1 and 2 series
g. 1 and 2 parallel
h. 1 and 3 parallel
i. 1 and 3 parallel

9. PE = physical exam MAM = mammography

p = 1% s(PE) = 52% s(MAM) = 65%

S(series) = 42% S(parallel) = 75%

a. If PE and MAM were independent, S(series) = 34%

S(parallel) = 83%

∴ the two tests are not statistically independent.

Also given: PVP(PE) = 20% PVP(MAM) = 25%

PVP(series) = 25% PVP(parallel) = 18%

b. In the tables below, TP is calculated from D and S, then FP and Pos are calculated using TP and PVP, the FN, TN, and Neg are obtained by subtraction. Then, c = TN/D

	+ PE	-
1	0.52	0.48
99	2.08	96.92
100	2.60	97.40

c(PE) = 97.9%

	+ MAM	-
1	0.65	0.35
99	1.95	97.05
100	2.60	97.40

c(MAM) = 98.0%

	+ SER -	
1	0.42	0.58
99	1.36	97.64
100	1.78	98.22

c(series) = 98.6%

	+ PAR -	
1	0.75	0.25
99	3.42	95.58
100	4.17	95.83

c(parallel) = 96.6%

c. Cost per cancer detected means cost of screening per TP detected

PE: $(100 \times 20 + 2.60 \times 400)/0.52 = \$5,846$

MAM: $(100 \times 40 + 2.60 \times 400)/0.65 = \$7,754$

Series: 100 women get PE, + ives get MAM, + ive of both go to biopsy:

$(100 \times 20 + 2.60 \times 40 + 1.78 \times 400)/0.42 = \$6,705$

Parallel (PE first): 100 women get PE, - ives get MAM, + ives in either go to biopsy:

$(100 \times 20 + 97.4 \times 40 + 4.17 \times 400)/0.75 = \$10,085$

Parallel (MAM first):

$(100 \times 40 + 97.4 \times 20 + 4.17 \times 400)/0.75 = \$10,155$

d. Cost of MAM = \$540, and add \$10,000 per FN

PE: $(100 \times 20 + 2.60 \times 400 + 0.48 \times 10,000)/0.52 = \$15,077$

MAM: $(100 \times 540 + 2.60 \times 400 + 0.35 \times 10,000)/0.65 = \$90,062$

Series:

$(100 \times 20 + 2.60 \times 540 + 1.78 \times 400 + 0.58 \times 10,000)/0.42 = \$23,610$

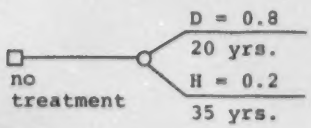
Parallel (PE first):

$(100 \times 20 + 97.4 \times 540 + 4.17 \times 400 + 0.25 \times 10,000)/0.75 = \$78,352$

Parallel (MAM first):

$(100 \times 540 + 97.4 \times 20 + 4.17 \times 400 + 0.25 \times 10,000)/0.75 = \$80,155.$

10.



$F_N = 0.8 \times 20 + 0.2 \times 35 = 23$ yrs.

Since $F_T = 31.35$ yrs. > F_N, treatment is indicated

11. $U_N(D) = 20$ yrs. $U_N(H) = 35$ yrs.

To calculate $U_T(D)$, $U_T(H)$, decision tree is folded back up to D and H branches:

$U_T(D) = 20 + 0.70 \times (0.01 \times 10 + 0.99 \times 15) + 0.30 \times 0.01 \times (-5) = 30.45$ yrs.

$U_T(H) = 35 + 0.01 \times (-5) = 34.95$ yrs.

Using (36): $A = \frac{30.45 - 20}{35 - 34.95} = 209$

Using (37): $p(\text{threshold}) = \frac{1}{1 + 209} = 0.00476.$

Since p > p(threshold), treatment is indicated.

12. Levels of disability associated with the seven outcomes:

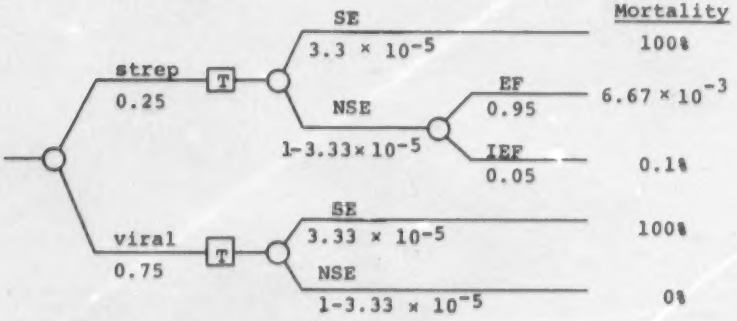
(1): 20% (4): 10%
(2): 5% (5) and (7): 0%
(3): 90% (6): 15%
Expected level of disability, using (31) = 3.34%

13. a. Alternative strategies:
(i) No culture, no treatment (send patient home)
(ii) No culture, give treatment from first day
(iii) Perform culture, treat if positive (second day)
(iv) Perform culture and start treatment, discontinue treatment if culture is negative
( ) is not a reasonable alternative
(iv) would be reasonable if dose dependent side effects of treatment were considered in problem.

b. Most important factor is mortality.

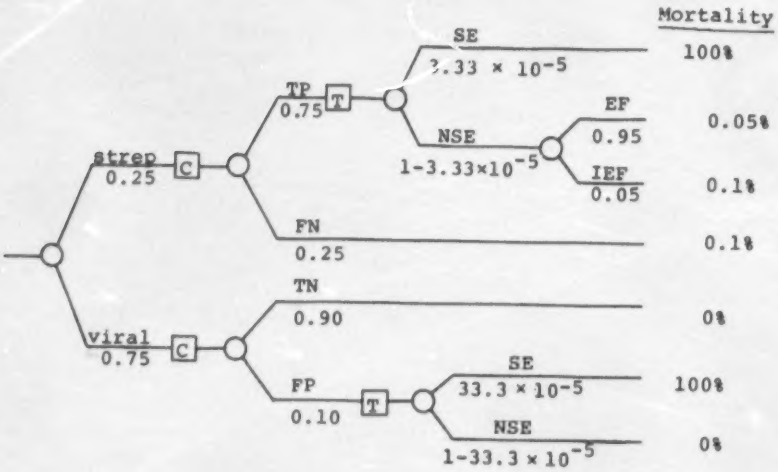
c. For reasons given above, only strategies (ii) and (iii) will be analyzed.

[C] = perform culture [T] = perform treatment
EF = treatment effective IEF = treatment ineffective
SE = side effect developed NSE = no side effect developed
Strategy (ii):



Expected mortality = 6.1 10^{-3}%.

Strategy (iii):



Probability of each outcome is calculated by multiplying probabilities of each branch leading to the outcome.

Expected mortality = 1.693×10^{-2}%.

On the basis of the analysis, strategy (ii) is preferred.

8. ANSWERS TO SAMPLE EXAM

1. Sensitivity and specificity are determined by applying the test to a population that is known to have the tested disease, and another that is known not to. Both s and c may depend on the particular population chosen, as well as on the demarcation line.

2. Prevalence is the fraction of a population that has a given disease at a given time, as determined by a particular criterion (e.g., clinical symptoms or a test). For an individual, prevalence is determined by the physician on the basis of age, sex, history, etc. Introduction of lead time by a new test will change the prevalence but not the incidence of disease.

3. Predictive values refer to the fraction of positive (or negative) results that are true. Since true and false results (for either positives or negatives) come from different populations, predictive values depend on prevalence. s and c refer to the fraction of sick and healthy populations, respectively, that show true results on the tests. Since each parameter refers to one population exclusively, s and c are independent of prevalence.

4. Experimental determination of predictive values requires the availability of a highly reliable method that can indicate the proportions of true results in the positive and negative findings of the test. The method can be another test with very high s and c, or a follow-up study of patients all the way to autopsy, if necessary.

5. All choices made by the physician are described as decision points, the branches corresponding to alternative choices (e.g., choice of diagnostic or therapeutic procedure). Probability nodes refer to situations where several alternatives are inherently present, each alternative being a probabilistic event outside the control of the physician (e.g., occurrence of false results, or side effects of therapy).

6. Both physician and patient contribute to subjective elements. The most important of these are: choice of diagnostic and therapeutic procedures, alternative diseases being considered, initial value of prevalence, cost/benefit factors, and weighting coefficients (if different factors are to be combined).

7. Advantage factor A is the ratio of two increments in utility: the first is the utility gained by treating a diseased subject as compared to not treating; the second is the utility gained by leaving a healthy subject untreated as compared to treating:

$$A = \frac{U_T(D) - U_N(D)}{U_N(H) - U_T(H)}.$$

If the treatment is an effective procedure, the numerator will be high. If the treatment is also safe, the denominator will be small. Under these conditions, A will be large and the minimum certainty of diagnosis to justify treatment (prevalence threshold) will be small.

8. PVP and PVN can be calculated from formulas, or the tables below:

a. p = 2%

	+	-	
D = 2	TP = 1.6	FN = 0.4	s = 80%
H = 98	FP = 4.9	N = 93.1	c = 95%
Tot = 100	Pos = 6.5	Neg = 93.5	

PVP = 0.246 PVN = 0.996

b. p = 1%

	+	-	
D = 1	TP = 0.8	FN = 0.2	s = 80%
H = 99	FP = 4.95	TN = 94.05	c = 95%
Tot = 100	Pos = 5.75	Neg = 94.25	

PVP = 0.139 PVN = 0.998

c. p = 10%

	+	-	
D = 10	TP = 8	FN = 2	s = 80%
H = 90	FP = 4.5	TN = 85.5	c = 95%
Tot = 100	Pos = 12.5	Neg = 87.5	

PVP = 0.64 PVN = 0.977

The test is best suited for individual diagnosis with relatively high prevalence. A negative result is still fairly reliable (97.7%), and a positive result (64% predictive value) can be confirmed by a culture. For screening purposes the FN is too high and a test with higher sensitivity is needed.

9. p = 10%, desired PVP = 80%, PVN = 99.9%, s = ?, c = ?

Using PVP and PVN formulas,

$$0.8 = \frac{0.1s}{0.1s + 0.9(1-c)} \qquad 0.999 = \frac{0.9c}{0.9c + 0.1(1-s)}$$

Rearranging,

$$s + 36c = 36 \quad \text{and} \quad 111s + c = 111.$$

Solving simultaneously,

$$s = 99.12\% \quad \text{and} \quad c = 97.25\%.$$

Sensitivity of IVP is acceptable but specificity is not high enough.

10. For screening purposes high sensitivity is desired, thus use tests in parallel.

$S = s_i$
$= 0.85 \times 0.55 \times 0.60$
$S = 0.28$
$C = 1 - (1-c_i)$
$= 1-0.25 \times 0.40 \times 0.20$
$C = 0.98$

$S = 1 - (1-s_i)$
$= 1-0.15 \times 0.45 \times 0.40$
$S = 0.973$
$C = c_i$
$= 0.75 \times 0.60 \times 0.80$
$C = 0.36$

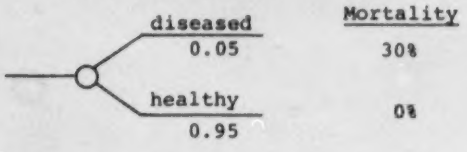
Series

	+	-	
D = 5	TP = 1.4	FN = 3.6	S = 28%
H = 95	FP = 1.9	TN = 93.1	C = 98%
Tot = 100	Pos = 3.3	Neg = 96.7	
	PVP = 0.4516	PVN = 0.9628	

Parallel

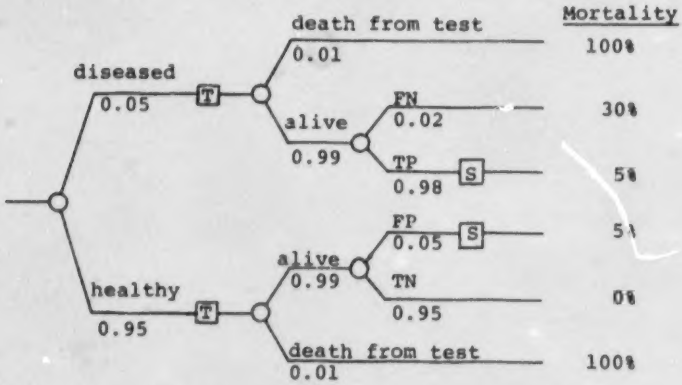
	+	-	
D = 5	TP = 4.865	FN = 0.135	S = 97.3%
H = 95	FP = 34.2	TN = 60.8	C = 36%
Tot = 100	Pos = 39.065	Neg = 60.935	
	PVP = 0.1245	PVN = 0.9978	

11. a. No testing or treatment:



Expected mortality = $0.05 \times 30 = 1.5\%$

b. Test (T) and surgery (S) if positive:



Expected
mortality = $(0.05 \times 0.01 + 0.95 \times 0.01) \times 100 + 0.05 \times 0.99 \times 0.02 \times 30 + (.05 \times 0.99 \times 0.98 + 0.95 \times 0.99 \times 0.05) \times 5 = 1.51\%$.

Expected mortalities in the two strategies are essentially equal. It would be best to look for alternative ways to substantiate the diagnosis.

c. Using mortality as a negative utility,

$U_N(H) = 0, \quad U_N(D) = -30.$

$U_T(D)$ and $U_T(H)$ are calculated by folding back the upper and lower three branches in (b), until D and H branches alone remain:

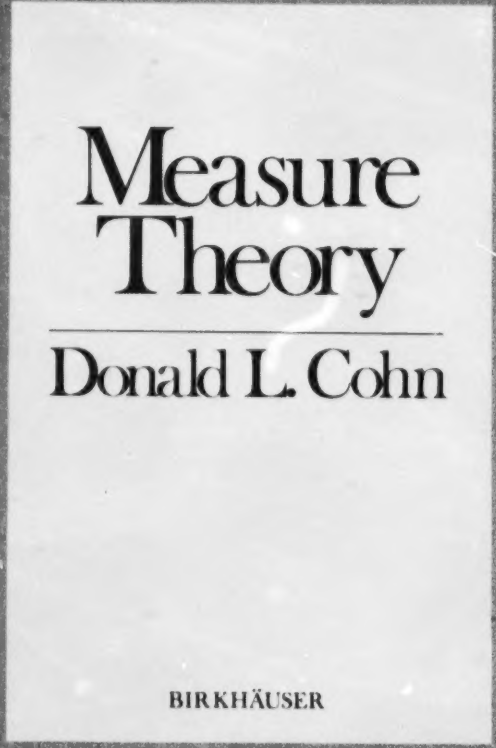
$U_T(D) = -(0.01 \times 100 + 0.99 \times 0.02 \times 30 + 0.99 \times 0.98 \times 5) = -6.445.$

$U_T(H) = -(0.01 \times 100 + 0.99 \times 0.05 \times 5) = -1.2475.$

$$A = \frac{-6.445 - (-30)}{0 - (-1.2475)} = 18.88$$

$$\text{Threshold } p = \frac{1}{1 + A} = 0.0503.$$

As expected from the comparison of the two strategies, the given level of certainty of diagnosis is barely equal to the required minimum.



Measure Theory

Donald L. Cohn

Clothbound 300 pages ISBN 3-7643-3003-1

May 1980

$19.95

An introduction to measure theory, emphasizing topics relevant and necessary for analysis and probability theory. Unique in its introductory, though comprehensive treatment of locally compact Hausdorff spaces, of the analytic and Borel subsets of Polish spaces, and of Haar measures on locally compact groups, this book provides the reader with the tools needed for study in several areas of current interest, harmonic analysis and probability theory, in particular.